

Das Universitätsranking im SPIEGEL der latenten Klassenanalyse

Tarnai, Christian; Grimm, Holger; John, Dirk; Watermann, Rainer

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Tarnai, C., Grimm, H., John, D., & Watermann, R. (1993). Das Universitätsranking im SPIEGEL der latenten Klassenanalyse. *ZA-Information / Zentralarchiv für Empirische Sozialforschung*, 33, 75-93. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-201517>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Das Universitätsranking im SPIEGEL der latenten Klassenanalyse

von Christian Tarnai, Holger Grimm, Dirk John und Rainer Watermann¹

Zusammenfassung

In dem vorliegenden Artikel wird die SPIEGEL-Studie des Jahres 1989 für das Fach Psychologie reanalysiert. Es wird gezeigt, unter welchen Voraussetzungen ein Universitäts-Ranking eine Beratungsgrundlage für Studierende bieten kann. Der Vorgehensweise der SPIEGEL-Studie bei der Erstellung der Rangreihe wird die Auswertung mittels latenter Klassenanalyse gegenübergestellt. Wesentliches Charakteristikum dieser Analyse ist die größtmögliche Berücksichtigung der Urteile einzelner Studierender. Ergebnis ist ein spezifisches Profil der beurteilten Studienbedingungen an jeder Universität, das Aussagen über Einheitlichkeit oder Uneinheitlichkeit der Beurteilung dieser Studienbedingungen zuläßt und sowohl negative als auch positive Bewertungen verdeutlicht. Dies kommt dem Informationsbedürfnis von Studierenden entgegen. Hierbei muß allerdings die Voraussetzung gemacht werden, daß die Urteile der Studierenden repräsentativ für die tatsächlichen Bedingungen am jeweiligen Hochschulort sind.

Abstract

The paper presents the reanalysis of the university-ranking published by the German news-magazine "Der SPIEGEL" in 1989. The Latent Class Analysis (LCA) is applied to the data in contrast to the original proceeding. The essential element of the LCA is to take into consideration all individual ratings of students as far as possible. The result is a specific profile of rated study-conditions for each university which meets the students' informational requirements. The principal assumption is that the ratings be representative according to the actual university's environment.

¹ Prof. Dr. Christian Tarnai, Justus-Liebig-Universität Gießen, FB Psychologie, Abt. Methodik, Otto-Behagel-Str. 10, D-35349 Gießen. Die Untersuchung wurde gemeinsam mit den Autoren Holger Grimm, Dirk John und Rainer Watermann an der Westf. Wilhelms-Universität Münster, Fachbereich Erziehungswissenschaft, Institut für Sozialpädagogik, Weiterbildung und Empirische Pädagogik, Georgskommende 33, D-48143 Münster, im Rahmen des Studienprojektes "Angewandte Methoden in der empirischen Bildungsforschung" durchgeführt.

1. Einleitung

Die *Neidhardt-SPIEGEL-Studie* (*SPIEGEL* 1990) hat in den letzten Jahren bei Studierenden, Hochschullehrern und all denen, die an der Ausbildungsqualität deutscher Hochschulen interessiert sind, für großes Aufsehen gesorgt und wohl nicht zu unrecht wurde die Art und Weise der Präsentation der Ergebnisse im *SPIEGEL-Spezial* kritisiert.

Erklärtes Ziel der vorgestellten Rangreihe der Universitäten ist es, wie dem Vorwort des *SPIEGEL-Spezials* zu entnehmen ist, "Abiturienten und Studenten bei der Wahl oder Einschätzung ihrer Hochschule [zu] beraten" (*SPIEGEL* 1990, S. 3). Das Beratungsangebot des *SPIEGELs* für die ratsuchenden Studierenden läßt sich auf Seite 12 und 13 des *SPIEGEL-Spezials* (1990) komprimiert in Form der "SPIEGEL-Rangliste der Universitäten" bewundern. Diese Rangliste basiert auf der Befragung von rund 6000 Studierenden, die ihre Studiensituation hinsichtlich Lehrveranstaltungen und Hochschullehrern für die 15 meist-studierten Hochschulfächer des alten Bundesgebietes einschätzen.

Die für die Auswertung wesentliche Vorgehensweise besteht darin, daß aus den Antworten auf die einzelnen Fragen über die Befragten eines Fachbereichs einer Hochschule (jeweils ca. N=12) Mittelwerte gebildet werden, welche die Grundlage für die weiteren Berechnungen darstellen, die zu der Rangreihe des Faches und über die Fächer hinweg zu einer Rangreihe der Hochschulen führen.

Die Kritik am Vorgehen und der Methode dieser Studie zur Erstellung einer Rangliste west-deutscher Hochschulen ist vielfältig (u.a. *Gräf* 1991; *Kreutz* 1992; *Lamnek* 1990; *Scheuch* 1990). Sie betrifft die Methode der Erhebung und die erwähnte Strategie der Auswertung, die durch fortlaufende Mittelwertbildung Einschätzungen nivelliert. Wenn man allen Kritikpunkten Rechnung trägt, kann den Antworten der rund 6000 Studierenden dann überhaupt noch ein Aussagewert zugeschrieben werden?

Neidhardt, wissenschaftlicher Berater der Studie, stellt hierzu fest: "[...] so zeigt sich schon jetzt, daß es mit den Instrumenten der empirischen Sozialforschung möglich ist, die bislang wirksamen "Gute-Ruf"-Vorstellungen [von Universitäten] zu überprüfen, zu korrigieren und zu qualifizieren. Was sich über universitäre Lehrverhältnisse bislang mit sehr starken Verzerrungen in einem bloßen Hören-Sagen gleichsam naturwüchsig bekannt machte und dann folgenreich war, läßt sich zuverlässiger transparent machen.

Die Kritiker dieses Verfahrens stehen nun unter dem Druck, zeigen zu müssen, ob und auf welche Weise dies noch besser geht, als es im ersten Anlauf gemacht wurde" (*SPIEGEL* 1990, S. 125).

In diesem Artikel wird, dem Aufruf *Neidhardts*, folgend, das Datenmaterial² beispielhaft für das Fach Psychologie für den Teilaspekt "Hochschullehrer" reanalysiert. Es wird der Versuch unternommen, die Frage zu beantworten, ob das SPIEGEL-Ranking in dieser Form überhaupt eine ernsthafte Beratungsgrundlage für Studierende bietet und inwieweit diesem Beratungsanspruch tatsächlich entsprochen wird. Es ist weiterhin zu klären, ob *Neidhardts* zentrale Annahme haltbar ist, daß die befragten Studierenden - in der Regel zwölf pro Fach - tatsächlich Experten für die Beurteilung von Studienbedingungen sind und daher ein gültiges Urteil über die gebotene Lehrqualität eines Faches an einer Universität abgeben können (z.B. *Neidhardt* 1990, S.120).

Zunächst werden die Stichprobe und die zu analysierende Skala "Hochschullehrer" beschrieben und in einer "klassischen" Analyse die im SPIEGEL verwendete Auswertungsstrategie nachgezeichnet und kritisiert, wobei in diesem Zusammenhang auch auf das aktuelle SPIEGEL-Ranking (*SPIEGEL* 1993) eingegangen wird. Im weiteren werden dann exemplarisch die Möglichkeiten der Latent Class Analyse (LCA) von *Lazarsfeld* (1950) vorgestellt, die im Gegensatz zur "klassischen Testtheorie" einen anderen Blick auf die Daten gewährt und neue Auswertungs-Perspektiven eröffnet. Die Latent Class Analyse (LCA) steht für eine Vielzahl von Modellen (*Lazarsfeld und Henry* 1968), die durch die Entwicklung von Rechenprogrammen in neuerer Zeit auch anwendbar geworden sind (vgl. *Formann* 1984; *Langeheine* 1988). Die Erweiterung der LCA durch *Rost* (1988a; 1988c) ermöglicht zudem die Integration von Modellen für ordinale Daten, die im Programm LA-CORD (*Rost* 1990) realisiert sind. Im Unterschied zu "klassischen" Verfahren benötigt die latente Klassenanalyse keinerlei metrische Qualität von Antworten. Die Beurteilungsähnlichkeit von Personen, die in der Neidhardt-SPIEGEL-Studie zwingende Voraussetzung ist, wird überprüft und im gleichen Schritt werden die Personen der Stichprobe nach dieser Beurteilungsähnlichkeit klassifiziert.

Für die beabsichtigte Auswertung ist es daher auch möglich, die Einschätzungen der einzelnen Studienbedingungen durch die Studierenden unabhängig von ihrer Zugehörigkeit zu einer Hochschule zu analysieren; das Gemeinsame der Studierenden ist also nicht wie bei der *Neidhardt*-SPIEGEL-Studie an die Zugehörigkeit an eine bestimmte Hochschule gebunden, sondern vielmehr wird diese Hochschulzugehörigkeit zum Kriterium für eine Urteilsübereinstimmung oder -abweichung. Dieser Analyseansatz ermöglicht einen wirklich qualitativen Vergleich von Fachbereichen und nicht bloß einen rein numerischen Vergleich von prozentuierten Mittelwerten, die pro Fach und Universität separat ermittelt werden.

2 Die Daten wurden vom Zentralarchiv für Empirische Sozialforschung, Universität zu Köln, aufbereitet, dokumentiert (ZA-Archiv-Nr. 1815 Rangliste der westdeutschen Hochschulen 1989) und zugänglich gemacht. Die Erhebung erfolgte im Auftrag der Zeitschrift DER SPIEGEL, die wissenschaftliche Leitung hatte Prof. Dr. *Friedhelm Neidhardt*, Berlin.

Daß die LCA ein dieser Fragestellung und den Daten angemessenes Verfahren ist, hat sich bereits durch Berechnungen für die Fächer Betriebswirtschaftslehre, Jura, Soziologie (*Tarnai* 1993) und Erziehungswissenschaft (*Tarnai et al.* 1993a) gezeigt. So läßt sich belegen, daß Klassifikationen der Studierenden auf der Grundlage ihrer Urteile möglich sind und diese mit der Hochschulzugehörigkeit zusammenhängen (*Tarnai* 1993). In der vorliegenden Arbeit wird der Vorteil der LCA im direkten Vergleich mit dem Vorgehen in der SPIEGEL-Studie herausgearbeitet.

2. Das SPIEGEL-Ranking

2.1 Stichprobe

Von den 5991 Studierenden der vorliegenden Untersuchung studieren 391 Personen Psychologie, was einem Gesamtstichprobenanteil von ca. 6,5 % entspricht. Die befragten Personen, von denen 55,8% (N=218) Studentinnen sind, verteilen sich auf 33 Universitäten.

2.2 Teilaspekt "Hochschullehrer"

Neidhardt bildet aus den Antworten von 18 Fragen, die die Studienbedingungen betreffen, zwei Teilaspekte (*SPIEGEL* 1990, S. 9). Ein Teilaspekt bezieht sich auf die Rahmenbedingungen (sieben Fragen), ein anderer auf das Engagement von Hochschullehrern (acht Fragen). Für alle Fächer ergeben sich somit drei fachspezifische Rangreihen, zwei Rangreihen für die Teilaspekte und eine, die alle 18 Fragen berücksichtigt und die eigentliche Rangliste des Faches darstellt. Im folgenden ist die Rangreihe für den Teilaspekt "Hochschullehrer" (Tabelle 1) wiedergegeben, auf den sich die folgenden Analysen beziehen. Die Beschreibung der Items in Tabelle 1 enthält Kurzfassungen der Originalfragen dieses Teilaspektes "Hochschullehrer" (vgl. *SPIEGEL* 1990, S. 8).

F2	Seminar-Sitzungen sind ausgefallen
F5	Persönliche Beratung durch Hochschullehrer möglich
F6	Hochschullehrer bereiten sich auf Lehrveranstaltungen vor
F7	Aufgaben in den Semesterferien werden gestellt
F8	Hochschullehrer erläutern Ergebnisse von Klausuren
F9	Hochschullehrer berücksichtigen Anregungen und Vorschläge
F10	Hochschullehrer besprechen Fragen laufender Forschung
F11	Hochschullehrer erklären fachliche Zusammenhänge verständlich

Tabelle 1: Kurzfassungen der Fragetexte für den Teilaspekt *Hochschullehrer*

Alle Fragen weisen sechs Antwortstufen auf, wobei allerdings die Benennung der Skalendepunkte nicht einheitlich ist. So werden Häufigkeitsurteile, die auf die Anzahl der Hochschullehrer bezogen sind (sehr wenige - sehr viele) mit Häufigkeitsurteilen, die sich auf Ereignisse beziehen (sehr selten - sehr häufig), miteinander vermischt.

Platz	Universität	Prozent	Platz	Universität	Prozent
1	Bielefeld	71,7	18	Münster	50,0
2	Würzburg	68,1	19	Konstanz	49,6
3	Bremen	67,1	20	Regensburg	49,4
4	Kiel	63,8	21	Berlin - FU	48,9
5	Oldenburg	60,4	22	Mannheim	48,6
6	Frankfurt	60,3	23	Mainz	47,5
7	Köln	56,5	24	Bochum	45,5
8	Osnabrück	55,9	25	Göttingen	44,6
9	Marburg	55,8	26	München	44,1
10	Wuppertal	54,8	27	Trier	42,5
11	Saarbrücken	52,5	28	Gießen	41,6
12	Bamberg	52,3	29	Freiburg	39,8
13	Tübingen	52,2	30	Erlangen-Nürnberg	37,2
14	Darmstadt	52,1	31	Hamburg	36,7
15	Bonn	51,9	32	Berlin - TU	35,6
16	Heidelberg	51,8	33	Braunschweig	33,3
17	Düsseldorf	51,5			

Tabelle 2: SPIEGEL-Rangliste für Psychologie: Hochschullehrer
Quelle: *SPIEGEL* (1990, S. 30)

Im methodischen Einleitungsartikel "Nicht Meinung, sondern Erfahrung" des *SPIEGEL-Spezials* (1990, S.6) wird zur Berechnung der verwendeten Ranking-Werte mitgeteilt: "Die Wertung eines einzelnen Fachbereichs entspricht dem Durchschnittswert aller durch die zwölf Studenten vergebenen Punkte. Dieses Ergebnis wiederum wird im Verhältnis zum theoretisch erreichbaren Höchstwert als Prozentzahl ausgewiesen. Das heißt: Einem Durchschnittswert von 6,0 entsprächen 100 Prozent, einem Durchschnittswert von 1,0 dagegen 0 Prozent. Miteinander verglichen und in eine gemeinsame Rangliste aufgenommen werden nur die Ergebnisse aller Fachbereiche desselben Fachs" (*SPIEGEL* 1990, S. 9). Legitimiert wird diese Berechnung durch die theoretische Annahme, daß die interviewten Studierenden Experten seien und zwölf von ihnen ausreichen, ein angemessenes Urteil über ihren Fachbereich abzugeben. *Neidhardt* begründet diese Annahme in dem Beitrag "Lob und Tadel sind befangen" (*SPIEGEL* 1990, S. 118), der sich mit ersten Reaktionen

auf das SPIEGEL-Ranking auseinandersetzt, folgendermaßen: "Zu den verbreitetsten Mißverständnissen gehört wohl, daß es sich [bei der Umfrage] um Meinungsforschung handle, bei der die Stichprobe der Befragten die Gesamtheit der Studenten ihres Fachbereichs zu repräsentieren habe. Darum ging es im vorliegenden Fall aber nicht. Gefragt war nicht nach Werten, Einstellungen oder Gefühlen der Studenten. Deshalb ist aus den Ergebnissen auch nicht (wie manche Kritiker meinen) herauszulesen, ob und in welchem Maße Studenten mit ihren Universitäten zufrieden sind. Die Aussagen, die mit der Umfrage fundiert werden sollten, bezogen sich nicht auf Studenten, sondern auf deren Studienbedingungen. Studenten fungieren in der Untersuchung nicht als Objekte, sondern als Informationsträger und Berichterstatter. Sie sind Experten, die über die von ihnen unmittelbar erfahrbaren Studienbedingungen Auskunft geben" (*SPIEGEL* 1990, S. 120). Weiter heißt es dann, daß sich auch niemand in anderen Bereichen nur auf das Urteil eines Experten, eines Berichterstatters verleiße. "Angesichts der Fragen, um die es bei der SPIEGEL-Umfrage geht, erschienen uns 12 pro Fachbereich als hinreichend. [...] Je mehr sich die Berichterstatter bestimmter Vorkommnisse oder Sachverhalte widersprechen, um so unsicherer bleibt man. Je mehr sich ihre Darstellung ähnelt, um so eher glaubt man, der Wahrheit nahezukommen. Konsensus gilt als eine Art Wahrheitsersatz" (*SPIEGEL* 1990, S. 120).

2.3 Berechnungsbeispiel

Für die Universität Düsseldorf, die genau den mittleren Rangplatz der Rangreihe "Hochschullehrer" einnimmt, zeigt ein einfacher Blick auf die Rohdatenmatrix (Tabelle 3) die Fragwürdigkeit der beschriebenen Expertentheorie. Zu sehen ist ein Konglomerat disparater Urteile; so berichtet z.B. Person p4 offensichtlich über einen anderen Fachbereich als Person p1. Dieser ungeschminkte Blick auf die Rohdaten wird durch die über alle Personen und Items gebildeten Randsummen Σ_i und die transformierten Meßwerte Σ_i/k ergänzt. In dieser Tabelle sind somit alle Basiswerte zur Berechnung des Ranglisten-Prozentwertes aufgeführt. Die transformierten Meßwerte müssen lediglich aufsummiert und die Summe durch die Anzahl der Personen geteilt werden. Der erzielte Mittelwert (\bar{x}_i) wird nach der oben genannten Vorschrift prozentuiert.

	F2	F5	F6	F7	F8	F9	F10	F11	Σ_i	$\Sigma_i/8$
p 1	6	6	6	3	6	6	3	6	42	5.25
p 2	4	4	4	1	1	2	3	2	21	2.63
p 3	2	4	3	3	2	4	4	3	25	3.13
p 4	1	3	3	1	1	4	5	3	21	2.63
p 5	5	2	4	4	1	1	3	2	22	2.75
p 6	2	5	5	1	2	4	2	4	25	3.13
p 7	6	6	6	2	2	3	3	2	30	3.75
p 8	4	2	3	5	2	3	3	2	24	3.00
p 9	3	4	3	5	4	2	3	3	27	3.38
p10	3	4	4	6	1	2	4	2	26	3.25
p11	6	6	5	1	5	4	5	5	37	4.63
p12	5	5	3	2	3	4	5	4	31	3.88
p13	6	5	6	4	5	5	5	5	41	5.13
										$\Sigma=372$ $\bar{x}=3.58$ (51,6%)
										$\Sigma/13=28.62$ $s=0.90$

Tabelle 3: Berechnungsbeispiel Universität Düsseldorf

Angenommen, dieses gängige Verfahren der Bildung eines additiven Index ist den vorliegenden Daten angemessen³, so ist es auch legitim den "wahren" Mittelwert (μ) aus den Daten zu berechnen, also das Vertrauensintervall dieses Mittelwertes zu bestimmen (vgl. Gräß 1991).

Nach der Formel $\mu = \bar{x} \pm z \cdot s/\sqrt{n}$ wird für die Universität Düsseldorf das Vertrauensintervall für eine Irrtumswahrscheinlichkeit von 5% berechnet. Es ergibt sich ein Mittelwerts-Intervall, das von 3,086 bis 4,068 reicht oder in SPIEGEL-Prozenten ausgedrückt ein Intervall zwischen 41,7% und 61,3% umfaßt. Die Universität Düsseldorf kann also genauso gut den Rangplatz 5 oder 28 einnehmen.

Diese Berechnung kann nun für alle Hochschulen durchgeführt werden, der Schwankungsbereich des Mittelwertes ist je Universität mehr oder minder stark ausgeprägt, jedoch insgesamt für eine Rangreihe fragwürdig.

³ Die Hauptkomponentenanalyse extrahiert für die Skala "Hochschullehrer" bezogen auf den Fachbereich Psychologie zwei Faktoren (Eigenwerte > 1). Die Bildung eines ungewichteten Mittelwertindex, der Eindimensionalität der Skala voraussetzen sollte, erscheint demnach fraglich.

2.4 Kritik und Konsequenzen: Die SPIEGEL-Rangreihe 1993

Obwohl im vorherigen Abschnitt nur eine einfache Berechnung beschrieben ist, kann, wenn die Ergebnisse anderer Autoren miteinbezogen werden (u.a. *Gräf* 1991; *Lamnek* 1990; *Scheuch* 1990), zur Auswertungsstrategie und zur Präsentation der gesamten Studie festgehalten werden: Die *Neidhardt*-SPIEGEL-Studie suggeriert in ihrer numerisch-graphischen Aufbereitung insbesondere im Mittelbereich der Rangreihen künstliche Unterschiede zwischen Universitäten und Fachbereichen, die so nicht zu rechtfertigen sind. Auch die Plazierungen der Universitäten in den oberen und unteren Bereichen der Rangreihen lassen nach Gevatter Zufall fragen, wenn zudem die Qualität der Datenerhebung mit Skepsis betrachtet wird.

Der Kritik ist Rechnung getragen worden: Die neue SPIEGEL-Rangreihe (*SPIEGEL* 1993) präsentiert sich in einem neuen methodischen und graphischen Gewand und auch die Qualität der Daten hat sich durch Kontrolle der Interviewer-Tätigkeit vermutlich verbessert. Zudem ist in dieser Studie die Anzahl von 12 befragten Studierenden pro Fachbereich auf 18 erhöht worden. Im wesentlichen entspricht der verwendete Fragebogen dem der ersten Studie, geändert hat sich angeblich das Antwortformat, das sich nun an der Schulnoten-Skala, die bekanntlich die Skalenwerte "sehr gut (1)" bis "ungenügend (6)" aufweist, orientiert. So heißt es wörtlich: "Die Beurteilung eines einzelnen Fachbereichs entspricht dem Durchschnittswert aller von den 18 Studenten vergebenen Noten"(!) (*SPIEGEL-Spezial* 1993, S. 9). Tatsächlich ist dem auf den Seiten 7 - 9 (*SPIEGEL-Spezial* 1993) abgedruckten Fragebogen zu entnehmen, daß die Studierenden auf den Großteil der Fragen mit den Häufigkeitseinschätzungen "sehr häufig (1)" oder nur "sehr selten (6)" antworten. Dieses Antwortformat ist bekannt, es findet in derselben Form im ersten Ranking (1990) Verwendung (s.o.), nur wurde die 6-Punkte-Skala dort noch nicht Schulnotenskala genannt und als solche auch nicht abgebildet. In der neuen Rangreihe des *SPIEGEL* (1993) sind nun also die Ranking-Werte in Form von Durchschnittsnoten dargestellt, die genauso wie oben beschrieben berechnet werden; auf eine zusätzliche Prozentuierung wurde allerdings verzichtet. Wirklich anders an dieser neuen Rangreihe ist, daß für die Berechnung der Rangwerte nun auch die Varianz der Mittelwerte berücksichtigt wird, was sich auch in der Darstellung der Rangliste widerspiegelt. "Nach den Regeln der Statistik konnten in jeder Fachrangliste drei Gruppen gebildet werden: Spitzengruppe, Mittelfeld und Schlußlichter. An den Gruppengrenzen kommt es zuweilen zu Überschneidungen. Fakultäten, die von den Studenten sehr einheitlich bewertet wurden, stehen dann trotz schlechterer Durchschnittsnote in einer höheren Gruppe als eher widersprüchlich benotete Fachbereiche" (*SPIEGEL* 1993, S. 84).

Was hat sich substantiell an der Methode und der Darstellung des SPIEGEL-Rankings 1993 im Vergleich zum SPIEGEL-Ranking 1989 geändert?

Wenig! Die Darstellung der Rangreihen in der Dreiteilung Spitzengruppe, Mittelfeld und Schlußlichter räumt zumindest einen Kritikpunkt des Rankings des Jahres 1989 aus. Höchst bedenklich ist jedoch, daß die verwendeten Antwortformate ("sehr viele - sehr wenige", "sehr selten - sehr häufig") in Form einer durchschnittlichen Schulnote "verkauft" werden. Inwieweit hier den "Regeln der Statistik" - wie immer sie aussehen mögen - gefolgt wurde, ist schwer nachzuvollziehen. Aber abgesehen davon gilt die grundsätzliche Kritik an der ersten Erhebung auch für die neue Rangreihe, denn auch hier besteht die für die Auswertung wesentliche Vorgehensweise darin, daß aus den Antworten auf die einzelnen Fragen über die Befragten eines Fachbereichs einer Hochschule Mittelwerte gebildet werden, welche dann die Grundlage für die weiteren Berechnungen darstellen, die zu der Rangreihe des Faches und über die Fächer hinweg zu einer Rangreihe der Hochschulen führen. Auch für die neue SPIEGEL-Studie werden die Einzelurteile durch die Mittelwertbildung nivelliert. Wie bereits in der Einleitung gesagt und wie es im folgenden demonstriert werden soll, ist es mit Hilfe der latenten Klassenanalyse möglich, die Einschätzungen der einzelnen Studienbedingungen durch die Studierenden unabhängig von ihrer Zugehörigkeit zu einer Hochschule zu analysieren; das Gemeinsame der Studierenden ist also nicht wie bei den beiden SPIEGEL-Studien an die Zugehörigkeit an eine bestimmte Hochschule gebunden, sondern vielmehr wird diese Hochschulzugehörigkeit zum Kriterium für eine Urteilsübereinstimmung oder -abweichung.

3. Latente Klassenanalyse

Die methodischen Grundlagen der vorgenommenen Analysen können im folgenden nur ansatzweise dargestellt werden. Es erscheint jedoch angebracht, zumindestens einige methodische Aspekte der latenten Klassenanalyse und deren Realisierung im Programm LACORD (*Rost* 1990) vorzustellen.

In Abgrenzung zur "klassischen Testtheorie" ist das wichtigste Merkmal probabilistischer Methoden, wie der LCA, die Möglichkeit, Annahmen über die Beziehung zwischen latenter und manifester Skala zu überprüfen, wobei dies auch die Überprüfung der Eigenschaften der manifesten Skala einschließt und nicht wie üblich per definitionem eine Verbindung zwischen der Skala der zu messenden Eigenschaft und der beobachteten Skala festgelegt wird. Allen probabilistischen Modellen ist gemeinsam, daß die Wahrscheinlichkeit einer Antwort als Funktion der latenten Eigenschaft aufgefaßt wird. Je nachdem, welche Funktion und Parameter angenommen werden, um die Antwortwahrscheinlichkeiten abzubilden, ergeben sich unterschiedliche probabilistische Modelle, wobei das bekannteste probabilistische Modell das Rasch-Modell ist, das mit allen seinen Erweiterungen von einem quantitativen latenten Kontinuum ausgeht (vgl. *Fischer* 1983). Ein Modell, das keine quantitative latente Personenvariable voraussetzt, ist die Latent Class Analyse (LCA) von *Lazarsfeld*

(1950), in der Personen qualitativ in Klassen entsprechend ihren Antwortwahrscheinlichkeiten eingeteilt werden. Rasch-Modell und LCA unterscheiden sich im wesentlichen nur im Skalenniveau der latenten Variablen, das im Rasch-Modell metrische und in der LCA nominale Qualität hat (vgl. **Rost und Strauß** 1992). In beiden Modellen wird die Wahrscheinlichkeit einer Person, genau eine Antwortkategorie bei einer Frage (Item) zu wählen, parametrisiert. Grundlage hierbei ist die mehrdimensionale Kontingenztafel, deren Abhängigkeiten durch Einführung einer latenten Variablen erklärt werden sollen. Dabei werden nicht nur die bivariaten, sondern alle in Frage kommenden Zusammenhänge höherer Ordnung berücksichtigt, wobei keine Zusammenhangsmaße, wie bei anderen Modellen latenter Variablen (z.B. Strukturgleichungsmodellen), benötigt werden⁴.

3.1 Berechnung

Ohne nun dezidiert auf die spezifische Art der Parametrisierung der Antwortwahrscheinlichkeiten der Personen über Items und Antwortkategorien hinweg einzugehen, die im Programm LACORD (**Rost** 1990) auf dem Konzept der Schwellenwahrscheinlichkeiten basieren (**Rost** 1988b; 1988c), errechnet sich das Ergebnis folgendermaßen: Das Programm LACORD ermittelt durch Maximum-Likelihood-Schätzungen anhand des Best Information Criterion (BIC) (vgl. **Bozdogan** 1987), das die Likelihood der Daten zur Anzahl der Modellparameter und dem Stichprobenumfang in Beziehung setzt (vgl. **Rost und Georg** 1991, S.58), das am besten zu den Daten passende Klassenmodell, in dem Personen ihren Antwortwahrscheinlichkeiten entsprechend optimal in Klassen angeordnet sind. Je nach Art der möglichen Repräsentation der Schwellendistanzen - der Abstände zwischen den Antwortstufen - können so vier verschiedene Modelle abgebildet werden. Sind die Parameter dieser Modelle nicht für alle latenten Klassen von Personen gleich, sondern jeweils unterschiedlich, ergeben sich vier weitere, sog. klassenspezifische Modelle (vgl. **Rost und Georg** 1991, S.63). Für diese Modelle ist es dann je nach Modelltyp und Gruppierung der Personen möglich, anhand der Schwellendistanzen die Qualität des Datenniveaus zu bestimmen, so daß gemeinsam mit der Klassifikation der Personen auch deren Gebrauch von Antwortstufen abgebildet werden kann.

Für die konkreten Analysen steht die Struktur der Urteile im Vordergrund. Auf dieser Grundlage werden die Personen klassifiziert. Dies geschieht anhand der jeweils höchsten Zuordnungswahrscheinlichkeit⁵. Diese Zuordnung ist notwendig, um in einem nächsten Schritt den Zusammenhang von Hochschulzugehörigkeit und Klassenzugehörigkeit herstellen zu können. Da es hierbei um die Personen der Stichprobe geht, werden bei der Darstel-

4 Detailliert sind die Grundlagen der Modelle, ihre Voraussetzungen und die Schätzmethoden der Parameter bei **Rost** (1988a) dargestellt.

5 Das Programm LACORD schätzt für jede Person ihre Zuordnungswahrscheinlichkeit pro Klasse (vgl. Fußnote 8).

lung der Urteilsstruktur anstelle der latenten Erwartungswerte⁶ die der Stichprobe verwendet; die numerischen Unterschiede sind dabei vernachlässigbar.

3.2 Analyseergebnisse

In die LACORD-Analyse gehen von 391 Personen 372 mit gültigen Werten (ca. 95%) in die Berechnungen ein. 19 Personen werden aufgrund fehlender Werte ausgeschlossen. Es zeigt sich, daß die Studierenden die acht Items der Skala Hochschullehrer (siehe Rohdatenmatrix für Düsseldorf) derart individuell beurteilen, daß sich 363 unterschiedliche Antwortmuster (Patterns) ergeben. Die LCA ist in der Lage, diese augenscheinliche Vielfalt der Urteile auf die in einer oder mehreren Klassen geltende gemeinsame (latente) Urteilsstruktur zurückzuführen. Für die vorliegenden Daten ergibt sich auf der Grundlage des BIC-Index eine 4-Klassenlösung.

BIC	Modell	Klassen	Likelihood	Parameter
9703.783	2	3	-4751.271	34
9693.742	2	4	-4719.615	43
9699.487	5	4	-4713.609	46
9693.811	5	4	-4695.974	51
9699.846	5	5	-4660.518	64
9707.852	6	4	-4655.643	67

Tabelle 4: Modellkennwerte der sechs besten Lösungen

Das dieser Lösung zugrunde liegende klassenunabhängige Modell (LACORD-Modell 2) der Schwellendistanzen gilt für alle vier Klassen in derselben Weise. Charakterisiert ist dieses Modell dadurch, wie Abbildung 1 zeigt, daß die Schwellendistanzen für alle Kategorien gleich groß, äquidistant, aber für die verschiedenen Items von unterschiedlicher Größe sind, was eine unterschiedliche Dispersion der Itemantworten bewirkt. Die Dispersion der Items, die durch die unterschiedlichen Längen der einzelnen Balken bzw. richtiger durch die halbe "Breite" einer Schwellendistanz repräsentiert wird, könnte für die "klassische" Testtheorie im weitesten Sinne mit der aus den Item-Analysen bekannten "Trennschärfe" verglichen werden. Der Skala "Hochschullehrer" kann aufgrund der Modell-Lösung Intervallskalenniveau attestiert werden, aber wie Abbildung 1 zeigt, müssen für jedes Item eigene Intervalle angenommen werden.

⁶ Im Gegensatz zur empirischen Zuordnungswahrscheinlichkeit sind die Erwartungswerte als die mittleren Antwortstufen zu verstehen, mit denen eine ideale Person in einer Klasse antwortet.

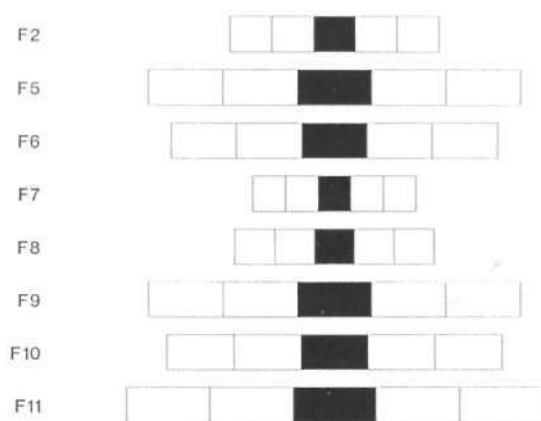


Abbildung 1: Schwellendistanzen Hochschullehrer

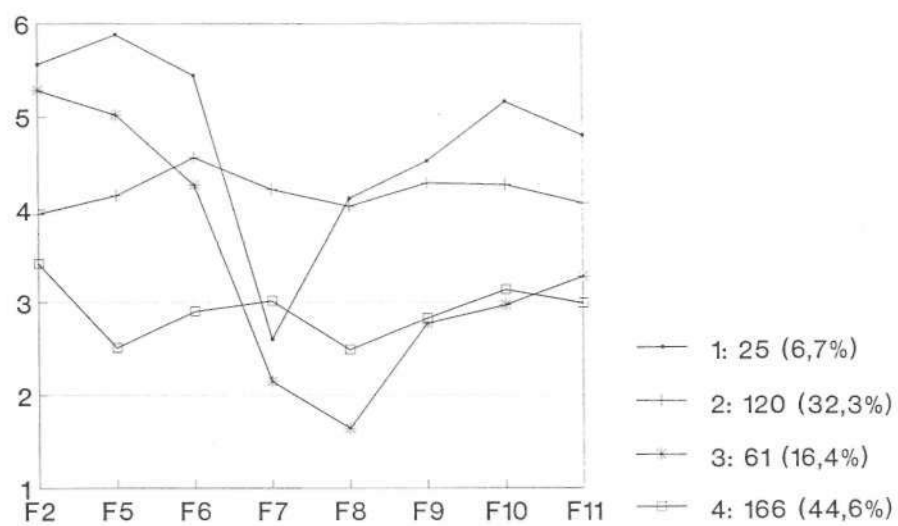


Abbildung 2: Erwartungswerte Hochschullehrer

Ein erster Blick auf die graphische LACORD-Lösung (Abbildung 2) zeigt, daß allein zwei Klassen - Klasse 2 und Klasse 4 - mehr als Dreiviertel der Personen in sich vereinigen. Klasse 2 zeigt über alle Items eine gleichmäßig positive Beurteilung, während Klasse 4 eine entsprechend negative Beurteilung widerspiegelt. Extremere Beurteilungen weisen Klasse 1 und Klasse 3 auf. Klasse 1 stellt mit 25 Personen, 6,7%, die kleinste Klasse dar. Sie weist für die Items F2 "Lehrveranstaltungen ausgefallen", F5 "Persönliche Beratung durch Hochschullehrer" und F6 "Hochschullehrer bereiten sich auf Lehrveranstaltungen vor" im Vergleich zu den anderen Klassen die positivsten Beurteilungen auf. Auch die andere "Extremklasse", Klasse 3 (16,4%), zeigt auf einem leicht niedrigeren Kurvenniveau eine ähnliche Beurteilungsstruktur. Eine deutliche Verschiebung des Beurteilungsniveaus in den negativen Bereich zeigen beide Klassen in dem Item F7 "Aufgaben in den Semesterferien". Die "Negativklasse" 4, mit 166 Personen (44,6%) die größte Klasse, nähert sich in ihrem Beurteilungsprofil in diesem Item dem Profil dieser beiden Klassen stark an, was für eine geringe Differenzierungskraft des Items F7 spricht (vgl. Abb. 1). Tatsächlich erscheint es auch inhaltlich fragwürdig, ob sogar "Experten" in der Lage sind, eine Frage nach den "Aufgaben in den Semesterferien" mit dem Antwortformat "1 = sehr selten", "6 = sehr häufig", objektiv zu bewerten.⁷ Ab dem Item F8 "Hochschullehrer erläutern Ergebnisse von Hausarbeiten und Klausuren" unterscheidet sich das Beurteilungsprofil der Klassen 1 und 3 deutlich; während Klasse 1 die Items F9 "Hochschullehrer berücksichtigen Anregungen und Vorschläge", F10 "Hochschullehrer besprechen Fragen laufender Forschung" und F11 "Hochschullehrer erklären fachliche Zusammenhänge" wieder positiv beurteilt, schätzen die Studierenden der Klasse 3 die Leistungen ihrer Hochschullehrer als weniger gut ein.

Wie bereits oben bemerkt bewegen sich die beiden Hauptklassen, Klasse 2 und Klasse 4, auf einem relativ konstanten Beurteilungsniveau. Während die größte Klasse, die "Negativklasse" mit 166 Personen (44,8%) ihre positivste Beurteilung im Item F2 findet, beurteilt die "Positivklasse", 120 Personen (32,3%), das Item F6 relativ am besten. Der näherungsweise parallele Profilverlauf dieser beiden Klassen weicht nur für F5 "Persönliche Beratung durch den Hochschullehrer" deutlich voneinander ab. Abschließend ist zu dieser Modell-Lösung festzustellen, daß die Profile der vier Klassen für die Studierenden der Psychologie eindeutige Beurteilungsunterschiede zeigen. Jeweils zwei Klassen (Klasse 1 und 3, Klasse 2 und 4) stehen eindeutig ordinal zueinander in Beziehung. Analysen für andere Fachbereiche (Soziologie, Jura, Betriebs- und Volkswirtschaft) haben für die von *Neidhardt* gebildeten Skalen ("Hochschullehrer", "Rahmenbedingungen") ähnlich mehrdimensionale Ergebnisse erbracht (vgl. *Tarnai* 1993). Wenn hinsichtlich der neuen SPIEGEL-Rangreihe behauptet wird, daß die für die Rangreihe verwendeten 18 Fragen (!) eine eindimensionale Skala (für alle Fachbereiche und Universitäten?) bilden und so die Verwendung des Mittelwertes als Gesamtbeurteilung legitim wäre (vgl. *SPIEGEL-Spezial* 1993, S. 154), sei in Unkenntnis der Rohdaten die Eindimensionalität der Skala schon jetzt

7 In der neuen SPIEGEL-Rangreihe wurde dieses Item F7 nicht mehr berücksichtigt.

bestritten. Die in der SPIEGEL-Studie eingesetzten Methoden mögen vielleicht auf Eindimensionalität hindeuten, aber eine latente Klassenanalyse würde zeigen, daß diese nicht besteht.

3.3 Klassifikation der Hochschulen

Nachdem die Beurteilungsstruktur der Studierenden des Faches Psychologie für den Bereich Hochschullehrer über alle Universitäten dargestellt worden ist, soll untersucht werden, in welchen Klassen Angehörige einzelner Universitäten lokalisierbar sind. Die Personen werden der Klasse zugeordnet, für die sie die höchste Wahrscheinlichkeit aufweisen⁸. Auf diese Weise kann nachvollzogen werden, inwieweit Studierende einer Universität als Experten im Sinne *Neidhardts* anzusehen sind. Es müßten nämlich alle Studierenden einer Hochschule in einer Klasse anzutreffen sein.

	Klasse 2	Klasse 4	Gesamt
	N	N	N
Bremen	12		12
Kiel	10	2	12
Tübingen	7	5	12
Regensburg	6	6	12
Heidelberg	4	5	9
Mainz	5	7	12
Bochum	4	7	11
Trier	3	8	11
Erlangen-Nürnberg	1	10	11
Hamburg	1	11	12
Braunschweig		12	12
Freiburg		12	12

Tabelle 5: Rangreihe der Hochschulen, die auf der durch die Klassen 2 und 4 repräsentierten Dimension abgebildet werden

Ausschließlich für die beiden zueinander ordinal angeordneten Klassen 2 und 4 ist es möglich, über ihre Angehörigen die Universitäten auf einer Dimension in eine Rangreihe zu bringen, weil alle ihre Studierenden eine einheitliche Beurteilungsgrundlage haben. Dies ist bei den ebenfalls zueinander ordinalen Klassen 1 und 3 nicht möglich, da bei keiner Hochschule ihre Studierenden ausschließlich in diesen beiden Klassen vertreten sind.

⁸ Die durchschnittliche Wahrscheinlichkeitszuordnung der Personen der Klasse 1 (N=25) beträgt 0,89, der Klasse 2 (N=120) 0,87, der Klasse 3 (N=61) 0,83, der Klasse 4 (N=166) 0,93.

Betrachtet man die Hochschulen, die nur in den Klassen 2 und 4 lokalisierbar sind, dann ist festzustellen, daß an der Universität Bremen die Studienbedingungen am positivsten gesehen werden. Alle ihre Studierenden befinden sich in der positiven Klasse 2. Auf der anderen Seite werden am negativsten die beiden Universitäten Braunschweig und Freiburg beurteilt. Von der Mehrheit der Studierenden positiv beurteilte Universitäten sind die Hochschulen Kiel und Tübingen. Von den insgesamt 12 Universitäten werden 8 mehrheitlich negativ eingeschätzt.

Die differenzierte Rangreihe, die sich aus den unterschiedlichen Urteilen der Angehörigen einer Universität ableiten läßt, ist in besonderem Maße von der Repräsentativität und der Anzahl der "Experten" abhängig. Entscheidend ist sicherlich die Repräsentativität der Urteile hinsichtlich der an einer Hochschule vorherrschenden Studienbedingungen. Eine Erhöhung der Anzahl von Studierenden, wie in der neuen SPIEGEL-Studie vorgenommen, macht nur dann Sinn, wenn die Repräsentativität dadurch verbessert wird.

	Klasse 1	Klasse 2	Klasse 3	Klasse 4	Gesamt
	N	N	N	N	N
Bamberg	2	3		7	12
Berlin-FU	1	1	1	6	9
Berlin-TU			3	9	12
Bielefeld	2	9	1		12
Bonn	1	4	4	2	11
Darmstadt		5	1	4	10
Düsseldorf	3	1	3	6	13
Frankfurt		8	1		9
Gießen			10	1	11
Göttingen			10	2	12
Köln	3	3	4	2	12
Konstanz		2	4	5	11
Mannheim	2		1	5	8
Marburg		7	1	4	12
München	1		4	6	11
Münster		2	4	6	12
Oldenburg		6	1	5	12
Osnabrück	5		2	4	11
Saarbrücken	2	1	2	5	10
Würzburg	3	8		1	12
Wuppertal		7	4	1	12

Tabelle 6: Klassifikation der Hochschulen, die nur qualitativ miteinander verglichen werden können

Für die beiden ebenfalls ordinalen Klassen 1 und 3 gibt es keine ausschließliche Zuordnung von Studierenden. Aus diesem Grunde können die verbleibenden 21 Hochschulen nur qualitativ miteinander verglichen werden, was gleichbedeutend mit dem Sachverhalt ist, daß keine Rangreihe über die Universitäten erstellt werden kann.

Dennoch können Unterschiede zwischen den Hochschulen deutlich nachvollzogen werden. Relativ homogen werden die Studienbedingungen an den Hochschulen beurteilt, deren Studierende in 2 Klassen zuordenbar sind. Es handelt sich hierbei um die Universitäten Berlin-TU, Gießen und Göttingen (Klasse 3 und 4) bzw. Frankfurt (Klasse 2 und 3). Die übrigen Universitäten, deren Angehörige in drei oder allen vier Klassen anzutreffen sind, lassen sich entsprechend der jeweiligen Anzahl ihrer Studierenden eindeutig qualitativ beschreiben. Möchte man eine Universität anhand des durch die Mehrheit ihrer Studierenden ausgedrückten Beurteilungsprofils charakterisieren, so kann das Kriterium, daß mehr als 50% in einer Klasse vereinigt sein sollen, als angemessen angesehen werden. Bei weniger Klassen ist ein strengeres Kriterium anzulegen (vgl. *Tarnai* 1993; *Tarnai et al.* 1993a).

Wenn man das vorliegende Ergebnis mit der SPIEGEL-Studie vergleicht, so sieht man deutlich, daß nur für wenige Universitäten die Voraussetzung für eine Mittelwertbildung und eine darauf aufbauende Rangreihe zutreffend ist. Nur für 12 Hochschulen ist eine Dimension Grundlage der Beurteilung, weshalb nur für sie das Kriterium des Expertentums als erfüllt anzusehen ist. Gegenüber der klassischen Vorgehensweise zeigt die latente Klassenanalyse, daß es möglich und sinnvoll ist, die Universitäten über die Urteile einzelner Personen eindeutig zu beschreiben. Dies erkennt man beispielsweise bei den drei Universitäten Erlangen-Nürnberg, Freiburg und Hamburg. Während die 12 Studierenden der Universität Freiburg sich in ihrem negativen Urteil einig sind (s. Tabelle 3), gibt es an den Universitäten Erlangen-Nürnberg und Hamburg jeweils eine Person, die diese Studienbedingungen positiv einschätzt. Das Gewicht der Einzelurteile wird dann deutlich, wenn keine eindeutige Mehrheit feststellbar ist, wie dies beispielsweise für Regensburg der Fall ist.

Der mit der Mittelwertbildung verbundene Informationsverlust kann anhand des Vergleichs der Universitäten Düsseldorf und Münster nochmals verdeutlicht werden (s. Tabelle 4). Beide Universitäten sind in der SPIEGEL-Rangreihe in benachbarter Position. Dennoch können Unterschiede anhand der Ergebnisse der latenten Klassenanalyse festgestellt werden. Die Mehrheit der Studierenden beider Universitäten urteilt negativ, während die übrigen Studierenden qualitative Unterschiede zwischen den Universitäten verdeutlichen. Aus den Ergebnissen ist ablesbar, daß die LCA in der Lage ist, gegebene Beurteilungsstrukturen vollständig zu erfassen, auch wenn sie qualitativ heterogen sind.

4. Schlußbemerkung

Die Ergebnisse der Analysen zeigen, daß die Urteile an einzelnen Hochschulen qualitativ und quantitativ auseinandergehen. Urteilsübereinstimmung stellt eher die Ausnahme dar. Selbst dort, wo die Mehrheit der 12 Studierenden einer Universität sich in ihrem Urteil einig ist, finden sich Partner mit qualitativ abweichendem Urteil. Somit ist zu schließen, daß ein Expertenurteil im Sinne des SPIEGELS ("Konsensus gilt als eine Art Wahrheitsersatz"; *SPIEGEL* 1990, S. 120) nicht sinnvoll ist. Wenn die Studienbedingungen an den meisten Hochschulen zu unterschiedlich sind, als daß sie überhaupt in derselben Weise wahrgenommen werden können, dann müssen Experten, die diese Studienbedingungen beurteilen, sich zwangsläufig "uneinig" sein.

Es wird deutlich, daß es nicht *die* besten oder schlechtesten Universitäten gibt. Es können für jede Universität Beurteilungsprofile herausgearbeitet werden. Dieses kommt dem Informationsbedürfnis von Studienanfängern entgegen, die über die Situation an den einzelnen Hochschulen Bescheid wissen wollen. Ein heterogenes Urteil zeigt an, daß heterogene Studienbedingungen vorliegen. Allerdings kann dieses Ergebnis nur dann als Beurteilungsgrundlage genommen werden, wenn die Leistungsprofile der einzelnen Hochschulen sich tatsächlich auf repräsentative Urteile ihrer Studierenden stützen und diese tatsächlich den angenommenen Objektbezug aufweisen. Ohne Berücksichtigung weiterer Daten kann nicht entschieden werden, ob die Urteile der Studierenden im angenommenen Sinne valide oder Ausdruck ihrer Persönlichkeit sind.

Literatur

Andrich, David, 1978:

A rating formulation for ordered response categories,
In: *Psychometrika*, 43, S. 561-573.

Andrich, David, 1982:

An Extension of the Rasch model for ratings providing both location and dispersion parameters,
In: *Psychometrika*, 47, S. 105-113.

Bozdogan, Hamparsum, 1987:

Model selection and Akaike's Information criterion (AIC): The general theory and its analytical extensions,
In: *Psychometrika*, 52, S. 345-370.

Der SPIEGEL.

Nr. 16 vom 19.04.1993.

Fischer, Gerhard H., 1983:

Neuere Testtheorie.

In *H. Feger & J. Bredenkamp* (Hrsg.), *Enzyklopädie der Psychologie*, Serie: Forschungsmethoden der Psychologie. Band 3: Messen und Testen (S. 604-692). Göttingen: Hogrefe.

Formann, Anton K., 1984:

Die Latent-Class-Analyse.

Weinheim: Beltz.

Gräf, Lorenz, 1991:

Fragwürdige Experten - Sekundäranalyse der SPIEGEL-Untersuchung zur Qualität westdeutscher Universitäten. *Soziologie* (Mitteilungsblatt der Deutschen Gesellschaft für Soziologie), 69-85.

Kreutz, Henrik, 1991:

Universitätsrangreihen und die Bewertung der Qualität von Hochschulen. Kritische Evaluation der SPIEGEL-rangreihe und der österreichischen Nachfolgestudie des Profil. Nürnberg: Lehrstuhl für Soziologie und Sozialanthropologie.

Lamnek, Siegrid, 1990:

Zur Professionalität der Studie: "Welche Uni ist die beste?"

Soziologie (Mitteilungsblatt der Deutschen Gesellschaft für Soziologie), 91-101.

Langeheine, Rolf, 1988:

New Developments in Latent Class Analysis.

In *R. Langeheine & J. Rost* (Eds.). *Latent Trait and Latent Class Models* (pp. 77-108). New York: Plenum.

Lazarsfeld, Paul F., Henry, Neil W., 1968:

Latent structure analysis.

Boston: Houghton Mifflin.

Lazarsfeld, Paul F., 1950:

The logical and mathematical foundation of latent structure analysis.

In *S. A. Stouffer, L. Guttman, E. A. Suchman, P. F. Lazarsfeld, S. A. Star, & J. A. Clausen* (Eds.), *Studies in social psychology in World War II. Vol. 4*. Princeton: Princeton University Press.

Masters, Geoffrey N., 1982:

A Rasch model for partial credit scoring.

Psychometrika, 47, 149-174.

Neidhardt, Friedhelm, 1990:

Lob und Tadel sind befangen. Über den Umgang mit dem SPIEGEL-Ranking.

In *SPIEGEL* (Hrsg.), *SPIEGEL-Spezial Nr. 1. Studieren heute. Welche Uni ist die beste?* (S. 118-125).

Hamburg: SPIEGEL Verlag.

Neidhardt, Friedhelm, 1991:

Kritik und Folgerungen - Am Beispiel des "SPIEGEL"-Ranking universitärer Lehrbedingungen.

Soziologie (Mitteilungsblatt der Deutschen Gesellschaft für Soziologie), 86-94.

Rost, Jürgen, 1988a:

Quantitative und qualitative probabilistische Testtheorie.

Bern: Huber.

Rost, Jürgen, 1988b:

Test theory with qualitative and quantitative latent variables.

In *R. Langeheine & J. Rost* (Eds.), *Latent Trait and Latent Class Models* (pp. 147-171). New York: Plenum.

Rost, Jürgen, 1988c:

Measuring Attitudes with a Threshold Model Drawing on a Traditional Scaling Concept.

Applied Psychological Measurement 12, 397-409.

Rost, Jürgen, 1990:

LACORD. Latent class analysis for ordinal variables.

A FORTRAN program. 2nd edition. Kiel: IPN.

Rost, Jürgen; Georg, Werner, 1991:

Alternative Skalierungsmöglichkeiten zur klassischen Testtheorie am Beispiel der Skala "Jugendzentrismus".

ZA-Information Nr. 28, 52-74.

Rost, Jürgen; Strauß Bernd, 1992:

Recent developments in psychometrics and test theory.

The German Journal of Psychology 16,91-119.

Scheuch, Erwin K., 1990:

Wie gut sind unsere Hochschulen?

Soziologie (Mitteilungsblatt der Deutschen Gesellschaft für Soziologie), 73-90.

SPIEGEL (Hrsg.), 1990:

Studieren heute. Welche Uni ist die beste?

SPIEGEL Spezial Nr. 1. Hamburg: SPIEGEL-Verlag.

SPIEGEL (Hrsg.), 1993:

Welche Uni ist die beste?

SPIEGEL Spezial Nr. 3. Hamburg: SPIEGEL-Verlag.

Tarnai, Christian, 1993:

Beurteilung der Studienbedingungen durch Studierende der Fächer Jura, Betriebswirtschaftslehre und Soziologie. Reanalyse der *Neidhardt-SPIEGEL-Studie* unter dem Blickwinkel individueller Urteile.

Sozialwissenschaftliche Forschungsdokumentationen. Münster: Institut für sozialwissenschaftliche Forschung e. V.

Tarnai, Christian; Grimm, Holger; John, Dirk; Watermann, Rainer, 1993a:

Studienbedingungen des Fachs Erziehungswissenschaft im SPIEGEL des Urteils von Studierenden.

Eine Detailanalyse der *Neidhardt-SPIEGEL-Studie* zur Universitätsrangreihe.

In *Ch. Tarnai* (Hrsg.), *Beiträge zur empirischen pädagogischen Forschung* (S. 145-165). Münster, New York: Waxmann.

Tarnai, Christian; Grimm, Holger; John, Dirk; Watermann, Rainer, 1993b:

Basis der Beurteilung von Studienbedingungen: Fach versus Studienort. Eine Reanalyse der

Neidhardt-SPIEGEL-Studie zur Universitätsrangreihe mittels Latent Class Analyse.

In *Ch. Tarnai* (Hrsg.), *Beiträge zur empirischen pädagogischen Forschung* (S. 166-178). Münster, New York: Waxmann.